UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 04-42
Statistics and Econometrics Series 11
September 2004

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

OUTLIER DETECTION IN MULTIVARIATE TIME SERIES VIA PROJECTION PURSUIT

Pedro Galeano*, Daniel Peña* and Ruey S. Tsay**

Abstract

This article uses Projection Pursuit methods to develop a procedure for detecting outliers in a multivariate time series. We show that testing for outliers in some projection directions could be more powerful than testing the multivariate series directly. The optimal directions for detecting outliers are found by numerical optimization of the kurtosis coefficient of the projected series. We propose an iterative procedure to detect and handle multiple outliers based on univariate search in these optimal directions. In contrast with the existing methods, the proposed procedure can identify outliers without pre-specifying a vector ARMA model for the data. The good performance of the proposed method is verified in a Monte Carlo study and in a real data analysis.

Keywords: Additive Outlier; Innovational Outlier; Level Change; Transitory Change; Projection Pursuit; Kurtosis coefficient.

Galeano, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail: pedro.galeano@uc3m.es. Peña, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail: daniel.pena@uc3m.es. Tsay, Graduate School of Business, University of Chicago, Chicago, IL 60637, USA, e-mail: ruey.tsay@gsb.uchicago.edu. The first two authors acknowledge financial support from BEC2000-0167, MCYT, Spain.

Outlier Detection in Multivariate Time Series Via Projection Pursuit

Pedro Galeano*, Daniel Peña* and Ruey S. Tsay**

* *Departamento de Estadística, Universidad Carlos III, Madrid, Spain*

** *Graduate School of Business, University of Chicago, Chicago, IL 60637, USA*

Abstract

This article uses Projection Pursuit methods to develop a procedure for detecting outliers in a multivariate time series. We show that testing for outliers in some projection directions could be more powerful than testing the multivariate series directly. The optimal directions for detecting outliers are found by numerical optimization of the kurtosis coefficient of the projected series. We propose an iterative procedure to detect and handle multiple outliers based on univariate search in these optimal directions. In contrast with the existing methods, the proposed procedure can identify outliers without pre-specifying a vector ARMA model for the data. The good performance of the proposed method is verified in a Monte Carlo study and in a real data analysis.

KEYWORDS: Additive Outlier; Innovational Outlier; Level Change; Transitory Change; Projection Pursuit; Kurtosis coefficient.

1 Introduction

Outlier detection in time series analysis is an important problem because the presence of even a few anomalous data can lead to model misspecification, biased parameter estimation and poor forecasts. Several detection methods have been proposed for univariate time series, including Fox (1972), Chang and Tiao (1983), Tsay (1986, 1988), Chang, Tiao and Chen (1988), Chen and Liu (1993), McCulloch and Tsay (1993, 1994), Le, Martin and Raftery (1996), Luceño (1998), Justel, Peña and Tsay (2000), Bianco et al (2001) and Sánchez and Peña (2003). Most of these methods are based on sequential detection procedures. For multivariate time series Tsay, Peña and Pankratz (2000) propose a detection method based on individual and joint likelihood ratio statistics.

Building adequate models for a vector time series is a difficult task, especially when the data are contaminated by outliers. In this paper, we propose a method to identify outliers without requiring initial specification of the multivariate model and is based on univariate outlier detection applied to some useful projections of the vector time series. The basic idea is simple: a multivariate outlier produces at least a univariate outlier in almost every projected series, and by detecting the univariate outliers we can identify the multivariate ones. We show that one can often identify better multivariate outliers by applying univariate test statistics to optimal projections than using multivariate statistics to the original series. We also show that in the presence of an outlier, the directions that maximize or minimize the kurtosis coefficient of the projected series include the direction of the outlier, that is, the direction that maximizes the ratio between the outlier size and the variance of the projected observations. We propose an iterative algorithm based on projections to clean the observed series from outliers.

This paper is organized as follows. In section 2 we introduce some notation and briefly review the multivariate outlier approach presented in Tsay et al. (2000). In section 3 we study properties of the univariate outliers introduced by multivariate outliers through projection and discuss some advantages of using projections to detect outliers. In section 4 we prove that the optimal directions to identify outliers can be obtained by maximizing or minimizing the kurtosis coefficient of the projected series. In section 5 we propose an outlier detection algorithm based on projections. We generalize the procedure to nonstationary time series in section 6 and investigate the performance of the proposed procedure in a Monte Carlo study in section 7. Finally, we apply the proposed method to a real data series in section 8.

2 Outliers in multivariate time series

Let $X_t = (X_{1t}, ..., X_{kt})'$ be a k-dimensional vector time series following the vector ARMA model

$$\Phi(B) X_t = C + \Theta(B) E_t, \quad t = 1, \cdots, n, \tag{1}$$

where B is the backshift operator such that $BX_t = X_{t-1}$, $\Phi(B) = I - \Phi_1 B - \cdots - \Phi_p B^p$ and $\Theta(B) = I - \Theta_1 B - \cdots - \Theta_q B^q$, are $k \times k$ matrix polynomials of finite degrees p and q, C is a k-dimensional constant vector, and $E_t = (E_{1t}, ..., E_{kt})'$ is a sequence of independent and identically distributed Gaussian random vectors with zero mean and positive-definite covariance matrix Σ. For the vector ARMA model in (1), we have the autoregressive representation $\Pi(B)X_t = C_\Pi + E_t$, where $\Pi(B) = \Theta(B)^{-1}\Phi(B) = I - \sum_{i=1}^{\infty}\Pi_i B^i$ and $C_\Pi = \Theta(1)^{-1}C$ is a vector of constants if X_t is invertible, and the moving-average representation $X_t = C_\Psi + \Psi(B)E_t$, where $\Phi(1)C_\Psi = C$ and $\Phi(B)\Psi(B) = \Theta(B)$ with $\Psi(B) = I + \sum_{i=1}^{\infty}\Psi_i B^i$.

Given an observed time series $Y = (Y_1', ..., Y_n')'$, where $Y_t = (Y_{1t}, ..., Y_{kt})'$, Tsay et al. (2000) generalize four types of univariate outliers to the vector case in a direct manner by using the representation

$$Y_t = X_t + \alpha(B)\, w I_t^{(h)}, \tag{2}$$

where $I_t^{(h)}$ is a dummy variable such that $I_h^{(h)} = 1$ and $I_t^{(h)} = 0$ if $t \neq h$, $w = (w_1, \cdots, w_k)'$ is the size of the outlier and X_t follows a vector ARMA model. The type of outlier is defined by the matrix polynomial $\alpha(B)$: if $\alpha(B) = \Psi(B)$, we have a multivariate innovational outlier (MIO); if $\alpha(B) = I$, we have a multivariate additive outlier (MAO); if $\alpha(B) = (I - B)^{-1}$, we have a multivariate level shift (MLS); and if $\alpha(B) = (I - \delta B)^{-1} I$, we have a multivariate temporary (or transitory) change (MTC), where $0 < \delta < 1$ is a constant. The effects of these outliers on the residuals are easily obtained when the parameters of the vector ARMA model for X_t are known. Using the observed series and the known parameters of the model for X_t, we obtain a series of residuals $\{A_t\}$ defined by

$$A_t = \Pi(B)Y_t - C_\Pi, \qquad t = 1, \ldots, n,$$

where $Y_t = X_t$ and $A_t = E_t$ for $t < h$. The relationship between the true white noise innovations E_t and the computed residuals A_t is given by

$$A_t = E_t + \Gamma(B)\, w I_t^{(h)}, \tag{3}$$

where $\Gamma(B) = \Pi(B)\alpha(B)$. Tsay et al. (2000) showed that when the model is known, the estimation of the size of a multivariate outlier of type i at time h is given by:

$$w_{i,h} = -\left(\sum_{j=0}^{n-h}\Gamma_j'\Sigma^{-1}\Gamma_j\right)^{-1}\left(\sum_{j=0}^{n-h}\Gamma_j'\Sigma^{-1}A_{h+j}\right), \quad i = I, A, L, T,$$

where $\Gamma_0 = -I$. The covariance matrix of this estimate is $\Sigma_{i,h} = \left(\sum_{j=0}^{n-h}\Gamma_j'\Sigma^{-1}\Gamma_j\right)^{-1}$. From (3), we have

$A_{h+j} = E_{h+j} - \Gamma_j w$, and can write

$$w_{i,h} = w - \left(\sum_{j=0}^{n-h} \Gamma_j' \Sigma^{-1} \Gamma_j \right)^{-1} \left(\sum_{j=0}^{n-h} \Gamma_j' \Sigma^{-1} E_{h+j} \right),$$

which implies that $\Sigma_{i,h}^{-1/2} w_{i,h}$ is distributed as $N\left(\Sigma_{i,h}^{-1/2} w, I\right)$. Thus, the multivariate test statistic

$$J_{i,h} = w_{i,h}{}' \Sigma_{i,h}^{-1} w_{i,h}, \qquad i = I, A, L, T \tag{4}$$

will be a non-central $\chi_k^2(\eta_i)$ with noncentrality parameters $\eta_i = w' \Sigma_{i,h}^{-1} w$, for $i = I, A, L, T$. In particular, under the null hypothesis $H_0 : w = 0$, the distribution of $J_{i,h}$ will be chi-squared with k degrees of freedom. A second statistic proposed by Tsay et al. (2000) is the maximum component statistic defined by

$$C_{i,h} = \max_{1 \leqslant j \leqslant k} \frac{|w_{j,i,h}|}{\sqrt{\sigma_{j,i,h}}}, \qquad i = I, A, L, T$$

where $w_{j,i,h}$ is the jth element of $w_{i,h}$ and $\sigma_{j,i,h}$ is the jth element of the main diagonal of $\Sigma_{i,h}$.

In practice the time index h of the outlier and the parameters of the model are unknown. The parameter matrices are then substituted by their estimates and the following overall test statistics are defined:

$$J_{\max}(i, h_i) = \max_{1 \leq h \leq n} J_{i,h}, \qquad C_{\max}(i, h_i^*) = \max_{1 \leq h \leq n} C_{i,h}, \qquad i = I, A, L, T \tag{5}$$

where h_i and h_i^* denote respectively the time index at which the maximum of the joint and maximum component statistics occur.

3 Outlier analysis through projections

In this section we explore the usefulness of projections of a vector time series for outlier detection. First, we study the relationship between the projected univariate models and the multivariate one. Second, we discuss some potential advantages of searching for outliers in the projected series.

3.1 Projections of a vector ARMA model

Let us study the properties of a univariate series obtained by the projection of a multivariate series that follows a vector ARMA model. It is well known that a non-zero linear combination of the components of the vector ARMA model in (1) follows a univariate ARMA model; see, for instance, Lütkepohl (1993). Let $x_t = v' X_t$. If X_t is a vector ARMA(p, q) process, then x_t follows an ARMA(p^*, q^*) model with $p^* \leqslant kp$ and $q^* \leqslant (k-1)p + q$. In particular, if X_t is a vector MA(q) series, then x_t is an MA(q^*) with $q^* \leqslant q$, and if X_t is a vector AR(p) process, then x_t follows an ARMA(p^*, q^*) model with $p^* \leqslant kp$ and $q^* \leqslant (k-1)p$. In general, the model of the univariate series is

$$\phi\left(B\right) x_t = c + \theta\left(B\right) e_t, \tag{6}$$

where $\phi\left(B\right) = \left|\Phi\left(B\right)\right|$, $c = v'\Phi\left(1\right)^* C$ and $v'\Omega\left(B\right) E_t = \theta\left(B\right) e_t$, where $\Phi\left(B\right)^*$ is the adjoint matrix of $\Phi(B)$, $\Omega(B) = \Phi(B)^*\Theta(B)$ and e_t is a scalar white noise process with variance σ_e^2. The values for $\theta\left(B\right)$ and σ_e^2 can be obtained using the algorithm proposed in Maravall and Mathis (1994), which always gives an invertible representation of the univariate process. The autoregressive representation of the univariate model (6) is $\pi(B)x_t = c_\pi + e_t$, where $c_\pi = \theta\left(1\right)^{-1} c$ and $\pi(B) = \theta\left(B\right)^{-1}\phi\left(B\right) = 1 - \sum_{i=1}^{\infty} \pi_i B^i$ and its moving-average representation is $x_t = c_\psi + \psi\left(B\right) e_t$, where $c_\psi = \phi\left(1\right)^{-1} c$ and $\psi\left(B\right) = \phi\left(B\right)^{-1}\theta\left(B\right) = 1 + \sum_{i=1}^{\infty} \psi_i B^i$.

When the observed series Y_t is affected by an outlier, as in (2), the projected series $y_t = v' Y_t$ satisfies $y_t = x_t + v'\alpha(B) w I_t^{(h)}$. Specifically, if Y_t has a multivariate additive outlier, the projected series is $y_t = x_t + \beta I_t^{(h)}$ so that it has an additive outlier of size $\beta = v'w$ at $t = h$ provided that $v'w \neq 0$. In the same way, the projected series of a vector process with a multivariate level shift of size w will have a level shift with size $\beta = v'w$ at time $t = h$. The same result also applies to temporary changes. Thus, for the three types of outliers mentioned above the following hypotheses are equivalent:

$$\begin{array}{l} H_0 : w = 0 \\ H_A : w \neq 0 \end{array} \Leftrightarrow \begin{array}{l} H_0^* : \beta = 0 \\ H_A^* : \beta \neq 0 \end{array} \qquad \forall v \in S^k - \{v \perp w\}$$

because $H_0 = \left\{\cap H_0^* : v \in S^k - \{v \perp w\}\right\}$, where $S^k = \left\{v \in R^k : v'v = 1\right\}$.

A multivariate innovative outlier produces a more complicated effect. It leads to a patch of consecutive outliers with sizes $v'w$, $v'\Psi_1 w, \cdots, v'\Psi_{n-h} w$, starting with time index $t = h$. Assuming that h is not close to n and because $\Psi_j \to 0$, the size of the outlier in the patch tends to zero. In the particular case that

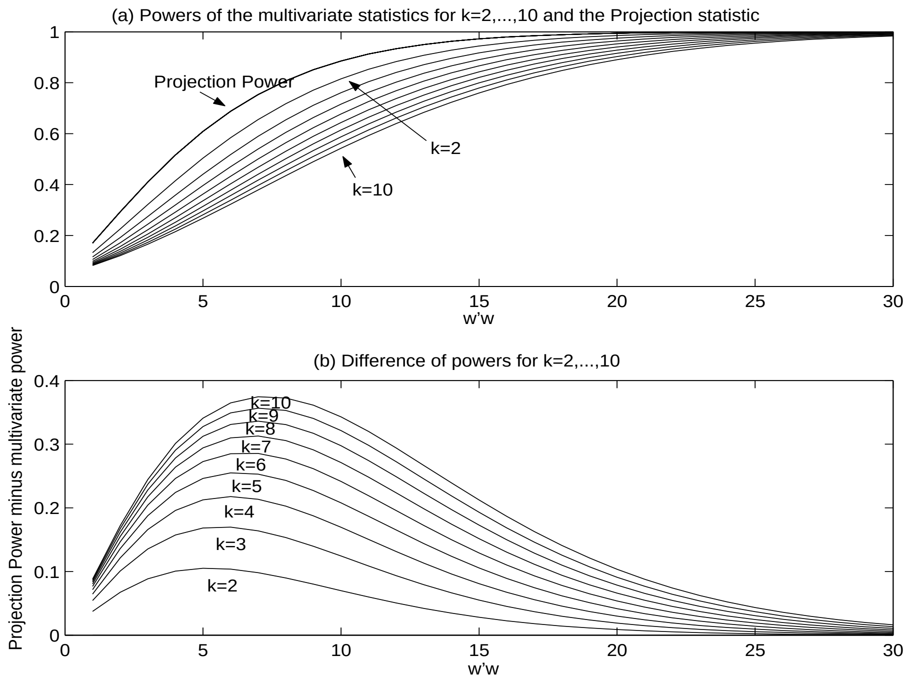

Figure 1: Powers of the Multivariate and the Projection statistics as a function of the outlier size: (a) Absolute Powers; (b) Difference of Powers.

$v'\Psi_i w = \psi_i v'w$, $\forall i = 1, \cdots, n-h$, then y_t has an innovational outlier at $t = h$ with size $\beta = v'w$. However, if $v'\Psi_i w = 0$, $i = 1, \ldots, n-h$, then y_t has an additive outlier at $t = h$ with size w, and if $v'\Psi_i w = v'w$, $i = 0, ..., n-h$, then y_t has a level shift at $t = h$ with size $\beta = v'w$. Therefore, the univariate series y_t obtained by the projection can be affected by an additive outlier, a patch of outliers or a level shift.

3.2 Some advantages of projection methods

The first advantage of using projections to search multivariate outliers is simplicity. By using univariate series we do not need to specify a multivariate model for the underlying series in outlier detection. Second, if the model parameters are known, a convenient projection direction will lead to test statistics that are more powerful than the multivariate ones. Third, as will be seen later in a Monte Carlo study, the same conclusion continues to hold when the parameters are estimated from the observed series.

To illustrate the second advantage, consider a k-dimensional time series Y_t generated from the vector ARMA model in (1) and affected by an MAO, MLS or MTC at $t = h$. Let V be the $k \times k$ matrix whose first column is $w/\|w\|$ and other columns are $k-1$ unit vectors orthogonal to w. The multivariate series $V'Y_t$ is affected by an outlier of size $(\|w\|, 0, \cdots, 0)'$ at time $t = h$. Notice that the outlier only affects the first component. Because the multivariate test statistic $J_{i,h}$ in (4) is invariant to linear transformations, its value is the same for both Y_t and $V'Y_t$ series. Thus, all the information concerning the outlier is in the first component of $V'Y_t$, which is the projection of the vector time series in the direction of the outlier. The remaining components of $V'Y_t$ are irrelevant for detecting the outlier. Moreover, because the test statistic $J_{i,h}$ is distributed as a non-central $\chi_k^2(\eta_i)$ with noncentrality parameter $\eta_i = w'\Sigma_{i,h}^{-1}w$ $(i = I,\, A,\, L,\, T)$, its power is given by $Pow\,(M) = \Pr\left(J_{i,h} > \chi^2_{k,\alpha}\right)$, where $\chi^2_{k,\alpha}$ is the 100α percentile of the chi-square distribution with k degrees of freedom. On the other hand, projecting the series Y_t on the direction v, we obtain a series y_t affected by an outlier at time $t = h$, and the univariate test statistic $j_{i,h} = \beta^2_{i,h}/\sigma^2_e$, where $\beta_{i,h}$ is the estimate of β, is distributed as a non-central $\chi^2_1(\widetilde{\eta}_i)$ with noncentrality parameter $\widetilde{\eta}_i = \beta^2/\sigma^2_{i,h}$, where $\beta = v'w$ and $\sigma^2_{i,h} = Var\,(\beta_{i,h})$. The power of this test statistic is $Pow(U) = \Pr\left(j_{i,h} > \chi^2_{1,\alpha}\right)$. Because the detection procedure we propose is affine equivariant, for simplicity we assume that Y_t is white noise and $\Sigma = I$. If $v = w/\,\|w\|$, then it is easy to see that for every w, $\eta_i = \widetilde{\eta}_i = w'w$ for $i = I$ and A, $\eta_L = \widetilde{\eta}_L = (n-h+1)w'w$ and $\eta_T = \widetilde{\eta}_T = \left(1-\delta^{2(n-h+1)}\right)/\left(1-\delta^2\right)w'w$. The powers, $Pow(U)$ and $Pow\,(M)$, and their differences $Pow(U) - Pow\,(M)$ for the case of an MAO are shown in Figure 1 for different values of $w'w$. The figure shows that the larger the number of components, the larger the advantage of the projection test over the multivariate one. When the size of the outlier increases both tests have power close to one and, hence, the difference goes to zero for large outliers. It will be seen in section 7 that for correlated series, the performance of both multivariate and projection test statistics depend on the model. We will also compare the power of both test statistics in section 7 via a simulation study. Finally, the same conclusion continues to hold when the parameters are estimated from the data.

4 Finding the Projection directions

The objective of Projection Pursuit algorithms is to find interesting features of high dimensional data in low dimensional spaces via projections. These projections are obtained by maximizing or minimizing an objective function named projection index, which depends on the data and the projection vector. The term *interesting projection* has often been associated with projections showing some unexpected structure such as clusters, outliers or non-linear relationships among the variables. It is commonly assumed that the most interesting projections are the farthest ones from normality. Some general reviews of Projection Pursuit techniques can be found in Huber (1985), Jones and Sibson (1987) and Posse (1995).

Peña and Prieto (2001a) showed that given two vector random variables having symmetric distributions with a common covariance matrix but different means, the direction that minimizes the kurtosis coefficient of the projection is the linear discriminant function, that is, the direction that produces the maximum separation between the projected means with respect to the variance of the projected distribution. These authors also propose a procedure for multivariate outlier detection based on projections that maximize or minimize the kurtosis coefficient of the projected data. Peña and Prieto (2001b) showed that these projected directions are also useful to identify clusters in multivariate data.

In this section we generalize the application of projections to multivariate time series analysis and define a maximum discrimination direction as the direction that maximizes the size of the univariate outlier, $v'w$, with respect to the variance of the projected series. We show that for multivariate additive outlier, level change and transitory change, the direction of the outlier is a direction of maximum discrimination and this direction can be obtained by finding the extreme of the kurtosis coefficient of the projected series. For a multivariate innovative outlier, we prove that the direction of the outlier is a maximum discrimination direction for the residual series and it can be obtained by projecting the residuals.

Let Y_t and A_t be the observed series and residuals in (2) and (3), respectively. For ease in presentation and without loss of generality, we assume $E(X_t) = 0$ and $\Sigma_X = \mathrm{Cov}(X_t) = I$, and define the deterministic variable,

$$R_t^{(h,n)} = \alpha\,(B)\,wI_t^{(h)} = wI_t^{(h)} - \alpha_1 wI_{t-1}^{(h)} - \cdots - \alpha_{n-h} wI_{t-(n-h)}^{(h)},$$

which contains two parameters, namely the time index h at which the outlier appears and the sample size n. Projecting Y_t on the direction v, we obtain $y_t = x_t + r_t^{(h,n)}$, where $r_t^{(h,n)} = v'R_t^{(h,n)}$. Let R_t and r_t be the coefficients of the variables $R_t^{(h,n)}$ and $r_t^{(h,n)}$ at the time index t, respectively, that is,

$$R_t = \begin{cases} 0 & t < h \\ -\alpha_{t-h} w & t \geq h, \end{cases} \qquad r_t = \begin{cases} 0 & t < h \\ -v'\alpha_{t-h} w & t \geq h, \end{cases}$$

where $\alpha_0 = -I$. Define $\overline{R} = \frac{1}{n}\sum_{t=1}^{n} R_t$ and $\overline{r} = \frac{1}{n}\sum_{t=1}^{n} r_t$, and let

$$E\left[\frac{1}{n}\sum_{t=1}^{n} Y_t\right] = \frac{1}{n}\left(I - \alpha_1 - \cdots - \alpha_{n-h}\right) w = \overline{R},$$

and

$$\Sigma_Y = E\left[\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \frac{1}{n}\sum_{t=1}^{n} Y_t\right)\left(Y_t - \frac{1}{n}\sum_{t=1}^{n} Y_t\right)'\right] = I + \Sigma_R,$$

where $\Sigma_R = \frac{1}{n}\sum_{t=1}^{n}(R_t - \overline{R})(R_t - \overline{R})'$. Using the results in Rao (1973, pg. 60), the maximum of $(v'w)^2 / (v'\Sigma_Y v)$ under the constraint $v'\Sigma_Y v = 1$ is $v = \Sigma_Y w$. In the cases of MAO, MLS and MTC, $\Sigma_Y = I + \beta_i w w'$, where β_i are given by

$$\beta_A = \frac{n-1}{n^2}, \quad \beta_L = \frac{n-h+1}{n}\left(\frac{h-1}{n}\right), \quad \beta_T = \frac{1}{n}\left[\left(\frac{1-\delta^{2(n-h+1)}}{1-\delta^2}\right) - \frac{1}{n}\left(\frac{1-\delta^{(n-h+1)}}{1-\delta}\right)^2\right]$$

and $v = (1 + \beta_i w'w)\, w$, implying that v is proportional to w. The same result holds in the MIO case for the maximum of $(v'w)^2 / (v'\Sigma_A v)$ under the constraint $v'\Sigma_A v = 1$, where Σ_A is the expected value of the covariance matrix of the innovations A_t.

Assuming that v verifies $v'\Sigma_Y v = 1$, the kurtosis coefficient of the series y_t is given by

$$\gamma_y(v) = E\left[\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \frac{1}{n}\sum_{l=1}^{n} y_l\right)^4\right].$$

To obtain the direction of the outlier, we prove next that w can be found by maximizing or minimizing the kurtosis coefficient $\gamma_y(v)$. These directions are solutions to the optimization problems:

$$\max_{v'\Sigma_Y v=1} \gamma_y(v) \qquad \text{and} \qquad \min_{v'\Sigma_Y v=1} \gamma_y(v). \tag{7}$$

To find the first-order conditions for (7), we need some preliminary results whose proofs are given in the appendix.

Lemma 1 *The kurtosis coefficient of y_t can be written as*

$$\gamma_y(v) = 3\left(v'\Sigma_Y v\right)^2 - 3\left(v'\Sigma_R v\right)^2 + \omega_r(v), \tag{8}$$

where $\omega_r(v) = \frac{1}{n}\sum_{t=1}^{n}(r_t - \overline{r})^4$.

Lemma 2 *The extreme directions of the kurtosis coefficient of y_t under the constraint $v'\Sigma_Y v = 1$ are given by the eigenvectors of the matrix*

$$\left[\sum_{t=1}^{n}\beta_t(v)\, B_t\right] v = \mu(v)\, v,$$

where $B_t = \left(R_t - \overline{R}\right)\left(R_t - \overline{R}\right)'$, $\beta_t(v) = (v'B_t v) - 3(v'\Sigma_R v) - \frac{\mu(v)}{n}$, and $\mu(v) = n\left(v'\Sigma_R v\right)^2(\gamma_r(v) - 3)$, where $\gamma_r(v)$ is the kurtosis coefficient of $r_t^{(h,n)}$. Moreover, the directions that maximize or minimize the *kurtosis coefficient are given by the eigenvectors linked to the largest and the smallest eigenvalues* $\mu(v)$*, respectively.*

The following result shows the usefulness of the directions that maximize or minimize the kurtosis coefficient of y_t.

Theorem 3 *Suppose* X_t *is a stationary vector ARMA*(p,q) *process and* $Y_t = X_t + \alpha(B) w I_t^{(h)}$.

1. *For a MAO, the kurtosis coefficient of* y_t *is maximized when* v *is proportional to* w *and it is minimized when* v *is orthogonal to* w.

2. *For a MTC, the kurtosis coefficient of* y_t *is maximized or minimized when* v *is proportional to* w *and it is minimized or maximized respectively when* v *is orthogonal to* w.

3. *For a MLS,*

   (a) *the kurtosis coefficient of* y_t *is minimized when* v *is proportional to* w *and it is maximized when* v *is orthogonal to* w *if*

$$h \in \left(1 + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right)n, 1 + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right)n\right),$$

   (b) *the kurtosis coefficient of* y_t *is maximized when* v *is proportional to* w *and it is minimized when* v *is orthogonal to* w *if*

$$h \notin \left(1 + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right)n, 1 + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right)n\right).$$

This theorem has two important implications. First, for a multivariate additive outlier, level shift or transitory change, one of the directions obtained by maximizing or minimizing the kurtosis coefficient is the direction of the outlier. Second, the directions are obtained without the information of the time index at which the outlier occurs.

Given the characteristics of innovational outliers, it is natural to think that the direction of the outlier can be easily obtained by focusing on the residual series. This is indeed the case.

Corollary 4 *If* X_t *is a stationary vector ARMA*(p,q) *process and* $Y_t = X_t + \Psi(B) w I_t^{(h)}$ *and* $A_t = E_t + w I_t^{(h)}$*, then the kurtosis coefficient of* $a_t = v' A_t$ *is maximized when* v *is proportional to* w *and it is minimized when* v *is orthogonal to* w.

On the other hand, it can be shown that the directions that produce the extreme values of the kurtosis coefficient in the presence of multiple outliers are linear combinations of the outlier sizes. Consequently, it would be of limited value in practice if one only considers the projections that maximize or minimize the kurtosis coefficient because of the potential problem of masking effects. To overcome such a difficulty, we propose to analyze a full set of $2k$ orthogonal directions consisting of (a) the direction that maximizes the kurtosis coefficient, (b) the direction that minimizes the kurtosis coefficient, and (c) two sets of $k-1$ orthogonal directions of (a) and (b). By doing so, if one of the outlier is masked in one direction, it can be revealed in one of the orthogonal directions. Furthermore, after detecting the outliers in the set of $2k$ orthogonal directions and cleaning their effects in the original series, we propose to iterate the analysis until no more outliers are detected.

5 Algorithms for outliers detection

We propose here a sequential procedure for outlier detection based on the directions of maximum discrimination. The procedure is divided into four steps: (1) obtain the projections of maximum discrimination; (2) search for outliers in the projected univariate time series; (3) all detected outliers in the univariate analysis are considered in a multivariate model framework and their effects are removed; (4) the procedure is applied again to the cleaned series until no more outliers are found. Note that in Step (2), the detection is carried out in two stages: first, level shifts are identified; second, innovative outliers, additive outliers and transitory changes are found. Finally, a vector model is identified for the cleaned time series and the outlier effects and model parameters are jointly estimated. The fitted model is refined if necessary, e.g. removing insignificant outliers if any.

5.1 Computation of the projection directions

We employ the procedure of Peña and Prieto (2001b) to construct the $2k$ projection directions of interest. For an observed vector series Y_t, our goal here is to solve the optimization problems in (7) and to obtain the orthogonal directions of the optimal projections. To this end, consider the procedure below:

1. Let $m = 1$ and $Z_t^{(m)} = Y_t$.
2. Define $\overline{Z}^{(m)} = \frac{1}{n}\sum_{t=1}^{n} Z_t^{(m)}$ and $\Sigma_Z^{(m)} = \frac{1}{n}\sum_{t=1}^{n}\left(Z_t^{(m)} - \overline{Z}^{(m)}\right)\left(Z_t^{(m)} - \overline{Z}^{(m)}\right)'$, and find v_m such

that

$$v_m = \underset{v_m' \Sigma_Z^{(m)} v_m = 1}{\arg\max} \frac{1}{n} \sum_{t=1}^{n} \left(v_m' Z_t^{(m)} - v_m' \overline{Z}^{(m)} \right)^4 . \tag{9}$$

3. If $m < k$, define

$$Z_t^{(m+1)} = \left(I - v_m v_m' \Sigma_Z^{(m)} \right) Z_t^{(m)},$$

that is, $Z_t^{(m+1)}$ is the projection of the observations in an orthogonal direction to v_m. Let $m = m + 1$. Otherwise, stop.

4. Repeat the same procedure to minimize the objective function in (9) to obtain another set of k directions; namely $v_{k+1}, \ldots, v_{2k}$.

A key step of the prior algorithm is to solve the optimization problem in (9). To this end, we employ a modified Newton method consisting of solving the system given by the first-order optimality conditions

$$\nabla \gamma_y (v) - 2\lambda \Sigma_Z^{(m)} v = 0$$
$$v' \Sigma_Z^{(m)} v - 1 = 0,$$

by means of linear approximations to these conditions. We refer interested readers to Peña and Prieto (2001b) for the technical details of the method. Note that the solutions obtained are local ones of the problems, but our simulations show that the adopted method works well. Another relevant issue is that the proposed procedure is affine equivariant, that is, the method selects equivalent directions for series modified by an affine transformation.

5.2 Searching for univariate outliers

The most commonly used tests for outlier detection in univariate time series are the likelihood ratio (LR) test statistics. Given a univariate time series y_t affected by an outlier at the time point $t = h$, the filtered series of residuals is defined by

$$a_t = e_t + \gamma(B) \beta I_t^{(h)},$$

where $\gamma(B) = 1 - \sum_{i=1}^{\infty} \gamma_i B^i$ such that $\gamma(B) = 1$ for an innovative outlier, $= \pi(B)$ for an additive outlier, $= (1 - B)^{-1} \pi(B)$ for a level shift and $= (1 - \delta B)^{-1} \pi(B)$ for a transitory change. The likelihood ratio test statistics for testing the hypothesis $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$ for each type of outlier are

$$\lambda_{i,h} = \frac{\beta_{i,h}}{\rho_{i,h}\sigma_e}, \qquad i = I, A, L, T$$

where $\rho_{i,h}^2 = \left(\sum_{j=0}^{n-h} \gamma_j^2\right)^{-1}$ with $\gamma_0 = -1$ and $\beta_{i,h} = -\rho_{i,h}^2 \left(\sum_{j=0}^{n-h} \gamma_j a_{h+j}\right)$ are the estimates of outlier sizes. Because $\lambda_{i,h}^2$ are the statistics $J_{i,h}$ in the case of $k = 1$, the distributions of $\lambda_{i,h}^2$ when the parameters are known are $\chi_1^2 (\widetilde{\eta}_i)$, where $\widetilde{\eta}_i = \left(\frac{\beta}{\rho_{i,h}\sigma_e}\right)^2$.

In practice the location h of the outlier and the parameters of the model are unknown. One uses the parameter estimates to define the overall test statistics

$$\Lambda(i, h_i) = \max_{1 \leq t \leq n} |\lambda_{i,t}|, \qquad i = I, A, L, T.$$

Using these statistics, Chang and Tiao (1983) propose an iterative algorithm for detecting innovational and additive outliers. Tsay (1988) generalizes the algorithm to detect level shifts and transitory changes. See Chen and Liu (1993) and Sánchez and Peña (2003) for additional extensions.

In this paper, we consider a different approach. There is substantial evidence that using the same critical values for all likelihood ratio test statistics can easily misidentify a level shift as an innovative outlier; see Balke (1993) and Sánchez and Peña (2003). The latter authors showed that the critical values for the likelihood ratio test statistic for detecting level shifts are different from those for testing additive or innovative outliers. Therefore, we propose to identify the level shifts in a series before checking for other types of outlier. To this end, it is necessary to develop a procedure that is capable of detecting level shifts in the presence of the other types of outliers. Carnero et al. (2003) show that the LR test for level shifts did not work well for financial time series and propose using a cusum test. Using the notation introduced in section 3.1, Bai (1994) shows that the cusum statistic

$$C_{h-1} = \frac{h-1}{\sqrt{n}\psi(1)\sigma_e} \left(\frac{1}{h-1} \sum_{t=1}^{h-1} y_t - \frac{1}{n} \sum_{t=1}^{n} y_t \right), \tag{10}$$

converges weakly to a standard Brownian Bridge on $[0, 1]$. Note that C_{h-1} is the statistic for testing a level shift at $t = h$. In practice, the term $\psi(1)\sigma_e$ is replaced by a consistent estimator and Bai (1994) recommends the following estimate

$$\widehat{\psi(1)\sigma_e} = \left[\widehat{\gamma(0)} + 2 \sum_{i=1}^{K} \left(1 - \frac{|i|}{K}\right) \widehat{\gamma(i)} \right]^{\frac{1}{2}},$$

where $\widehat{\gamma(h)} = \mathrm{Cov}(x_t, x_{t-h})$ and K is a quantity such that $K \longrightarrow \infty$ and $K/n \rightarrow 0$ as $n \rightarrow \infty$; see Priestley (1981). The statistic $\max_{1 \leq t \leq n} |C_t|$ under the assumption of no level shifts in the sample is asymptotically distributed as the supremum of the absolute value of a Brownian Bridge with cumulative distribution function (Billingsley, 1968),

$$F(x) = 1 + 2 \sum_{i=1}^{\infty} (-1)^i e^{-2i^2 x^2}, \qquad x > 0,$$

and Bai (1994) shows the consistency of this statistic for detecting the change point.

The cusum statistic (10) has several advantages over the LR statistic for detecting level shifts. First, the asymptotic distribution is independent of the error distribution so the Gaussian assumption is not required. Second, it is not necessary to specify the order of the ARMA model, which can be difficult under the presence of level shifts. Third, as shown in section 7, this statistic seems to be more powerful than the LR in all the models considered. Fourth, the statistic (10) seems to be robust to the presence of other outliers whereas the LR test statistic is not.

5.2.1 Level shift detection

Given the $2k$ projected univariate series $y_{t,j} = v_j' Y_t$ for $j = 1, \cdots, 2k$, we propose an iterative procedure to identify level shifts based on the algorithm proposed in Inclán and Tiao (1994) for detecting variance changes and Carnero et al. (2003) for identifying level shifts in a white-noise financial time series. The algorithm divides the series into pieces after detecting a level shift, and proceeds as follows:

1. Let $t_1 = 1$.

2. Obtain

$$D_L = \max_{1 \leq i \leq 2k} \max_{1 \leq t \leq n} \left| C_t^i \right|, \tag{11}$$

   where C_t^i is given by (10) for $t = 1, \ldots, n$ and the $2k$ series. Obtain

$$(t_{\max}, i_{\max}) = \arg\max_{1 \leq i \leq 2k} \arg\max_{1 \leq t \leq n} \left| C_t^i \right|.$$

   If $D_L > D_{L,\alpha}$, then there is a possible level shift at $t = t_{\max} + 1$, where $D_{L,\alpha}$ is the critical value for the significant level α. If $D_L < D_{L,\alpha}$, then there is no level shift in the series.

3.a Define $t_2 = t_{\max}$ of Step 2, and obtain

$$(t_{\max}, i_{\max}) = \underset{1\le i\le 2k}{\arg\max}\ \underset{1\le t\le t_2}{\arg\max} \left|C_t^i\right|.$$

If $D_L > D_{L,\alpha}$, then we redefine $t_2 = t_{\max}$ and repeat Step 3.a until $D_L < D_{L,\alpha}$. Define $t_{first} = t_2$ where t_2 is the last value that attains the maximum of the cusum statistics and is larger than $D_{L,\alpha}$. The point $t_{first} + 1$ is the first time point with a possible level shift.

3.b We repeat a similar search in the interval $t_2 \le t \le n$, where t_2 is the point $t_{\max}$ obtained in Step 2. Furthermore, define $t_1 = t_{\max} + 1$, where

$$(t_{\max}, i_{\max}) = \underset{1\le i\le 2k}{\arg\max}\ \underset{t_1\le t\le n}{\arg\max} \left|C_t^i\right|,$$

and repeat the process until $D_L < D_{L,\alpha}$. Let $t_{last} = t_1 - 1$, where t_1 is the last value that attains the maximum of the cusum statistics and is larger than $D_{L,\alpha}$.

3.c If $|t_{last} - t_{first}| < H$, where H is an integer defining the smallest interval between two level shifts, there is just a level shift and the algorithm finishes. If not, keep both values as possible change points and repeat the Steps 2 and 3 for $t_1 = t_{first}$ and $n = t_{last}$ until no more possible change points are detected. Then, go to Step 4.

4. Define a vector $h^L = \left(h_1^L, \ldots, h_{r_L}^L\right)$ where $h_1^L = 1$, $h_{r_L}^L = n$ and $h_2^L, \ldots, h_{r_L-1}^L$ are the change points detected in Steps 2 and 3 in increasing order. Obtain the statistic D_L in each sub-intervals $\left(h_i^L, h_{i+2}^L\right)$ and check its significance. If a D_L is not statistically significant, eliminate the corresponding possible change point. Repeat Step 4 until the number of possible change points remains unchanged and the time indexes found do not differ from those of the previous iteration for two time periods. Removing the points $h_1^L = 1$ and $h_{r_L}^L = n$ from the final vector of time indexes, we obtain the time points of level shifts by adding one to those remain in the final vector.

Some comments on the procedure are in order. First, one can rewrite the statistic (11) as

$$D_L = \max_{1\le i\le 2k}\ \max_{1\le t\le n} \left|C_t^i\right| = \max_{j=1,2}\ \max_{1\le i_j\le k}\ \max_{1\le t\le n} \left|C_t^{i_j}\right|,$$

where j is 1 for the k directions of the maximum and is 2 for the k directions of the minimum. Thus, D_L is the maximum of two dependent random variables. This dependence makes the distribution of D_L

intractable. We obtain critical values for different significant levels via simulation in the next section.

Second, consider the number H in Steps 3.c and 4. From the definition, the test statistics (10) are highly correlated for h close to each other. Thus, consecutive large values of C_{h-1} might be caused by a single level shift. To avoid over detection, we do not allow two level shifts to be too close. In the simulations and real data example, we chose H to be the number of estimated parameters plus one, that is

$$H = k(p+q+1) + \frac{k(k+1)}{2} + 1$$

and found it works well.

Let $\{h_1^L, \ldots, h_{r_L}^L\}$ be the time indexes of r_L detected level shifts. To remove the impacts of level shifts, we fit the following model

$$\left(I - \Pi_1 B - \ldots - \Pi_{\widehat{p}} B^{\widehat{p}}\right) Y_t^* = A_t^*, \tag{12}$$

where $Y_t^* = Y_t - \sum_{i=1}^{r_L} w_i S_t^{(h_i^L)}$, and the order $\widehat{p}$ is chosen such that

$$\widehat{p} = \underset{0 \le p \le p_{\max}}{\arg\min}\, AIC(p) = \underset{0 \le p \le p_{\max}}{\arg\min} \left\{ \log\left|\widehat{\Sigma}_p\right| + 2\frac{k^2 p}{n} \right\},$$

where $\widehat{\Sigma}_p = \frac{1}{n-2p-1}\sum_{t=p+1}^{n} A_t^* A_t^{*\prime}$ and $p_{\max}$ is a prespecified upper bound. If some of the effects are not significant, we remove the least significant one from the model (12) and re-estimate the effects of the remaining $r_L - 1$ level shifts. This process is repeated until all the level shifts are significant.

5.2.2 Algorithms for outliers detection

Using the level-shift adjusted series, we propose a procedure to detect additive outliers, transitory changes and innovative outliers as follows:

1. Obtain the $2k$ directions that maximize or minimize the kurtosis coefficient of the projected series of Y_t^* and their orthogonal directions. Denote the projected series by $y_{t,j}$ for $j = 1, \cdots, 2k$. Obtain also another $2k$ directions that maximize or minimize the kurtosis coefficient of the projected series from the residual A_t^* and their orthogonal directions. Denote the projected series by $a_{t,1}, \ldots, a_{t,2k}$.

2. For each univariate series $y_{t,i}$, we fit an autoregressive model with order selected by the Akaike information criterion (AIC). For $t = 1, \ldots, n$, compute the test statistics, $\lambda_{A,t}^i$ and $\lambda_{T,t}^i$, $i = 1, \ldots, 2k$, using the parameter estimates of the autoregression. Obtain the maximum of the statistics $\left|\lambda_{A,t}^i\right|$ and

$\left|\lambda^i_{T,t}\right|$ for each series, and then, the maxima across the series. On the other hand, for each univariate residual series $a_{t,i}$, compute the test statistics $\left|\lambda^i_{I,t}\right|$, where $i = 1, \dots, 2k$, and obtain the maximum of the statistics $\left|\lambda^i_{I,t}\right|$ over all time points and across series. Thus, we obtain

$$\Lambda_A = \max_{1\leq i\leq 2k} \max_{1\leq t\leq n} \left|\lambda^i_{A,t}\right|, \quad \Lambda_T = \max_{1\leq i\leq 2k} \max_{1\leq t\leq n} \left|\lambda^i_{T,t}\right|, \quad \Lambda_I = \max_{1\leq i\leq 2k} \max_{1\leq t\leq n} \left|\lambda^i_{I,t}\right|. \tag{13}$$

3. Let $\Lambda_{A,\alpha}$, $\Lambda_{T,\alpha}$ and $\Lambda_{I,\alpha}$ be the critical values for a predetermined significant level α. There are three possibilities:

   (a) If $\Lambda_j < \Lambda_{j,\alpha}$, $j = I$, A, T, no outliers are found and go to Step 4.

   (b) If $\Lambda_j > \Lambda_{j,\alpha}$ for only one j, where $j = A, T, I$, we identify an outlier of type j and remove its effect using the multivariate parameter estimates.

   (c) If $\Lambda_j > \Lambda_{j,\alpha}$ for more than one j, we identify the most significant outlier and remove its effect using the multivariate parameter estimates.

   We repeat Steps 1, 2, and 3 until no more outliers are detected.

4. Let $\{h^A_1, \dots, h^A_{r_A}\}$, $\{h^T_1, \dots, h^T_{r_T}\}$ and $\{h^I_1, \dots, h^I_{r_I}\}$ be the time indexes of the r_A, r_T and r_I detected additive outliers, transitory changes and innovative outliers, respectively. We estimate jointly the model parameters and the detected outliers for the series Y^*_t:

$$\left(I - \Pi_1 B - \dots - \Pi_{\widehat{p}} B^{\widehat{p}}\right) Y^{**}_t = A^{**}_t,$$

   where

$$Y^{**}_t = Y^*_t - \sum_{i_A=1}^{r_A} w_{i_A} I_t^{\left(h^A_{i_A}\right)} - \sum_{i_T=1}^{r_T} \frac{w_{i_T}}{1-\delta B} I_t^{\left(h^T_{i_T}\right)}, \qquad A^{**}_t = A^*_t - \sum_{i_I=1}^{r_I} w_{i_I} I_t^{\left(h^I_{i_I}\right)}.$$

   If some of the effects are not significant, we remove the least significant outlier. This process is repeated until all the outliers are significant.

The critical values for the statistics $\lambda^i_{A,t}$, $\lambda^i_{T,t}$ and $\lambda^i_{I,t}$ are obtained via simulation. In section 7, several critical values for different models, number of components and sample sizes are given.

5.3 Final joint estimation of parameters, level shifts and outliers

By now, we have a number of detected level shifts and outliers, and proceed to perform a joint estimation of the model parameters, the level shifts and the outliers using the equation

$$\left(I - \Pi_1 B - \ldots - \Pi_{\widehat{p}} B^{\widehat{p}}\right) Z_t = D_t,$$

where

$$Z_t = Y_t - \sum_{i_L=1}^{r_L} w_{i_L} S_t^{\left(h_{i_L}^L\right)} - \sum_{i_A=1}^{r_A} w_{i_A} I_t^{\left(h_{i_A}^A\right)} - \sum_{i_T=1}^{r_T} \frac{w_{i_T}}{1-\delta B} I_t^{\left(h_{i_T}^T\right)}, \qquad D_t = A_t - \sum_{i_I=1}^{r_I} w_{i_I} I_t^{\left(h_{i_I}^I\right)},$$

and $\left\{h_1^L, \ldots, h_{r_L}^L\right\}$, $\left\{h_1^A, \ldots, h_{r_A}^A\right\}$, $\left\{h_1^T, \ldots, h_{r_T}^T\right\}$ and $\left\{h_1^I, \ldots, h_{r_I}^I\right\}$ are the time indexes of the r_L, r_A, r_T and r_I detected level shifts, additive outliers, transitory changes and innovative outliers, respectively. If some effect (outlier or level shift) is found not significant at a given level, we remove the least significant effect and repeat the joint estimation until all the effects are significant.

6 The nonstationary case

In this section we study the case that the time series is unit-root nonstationary. Assume $X_t \sim I(d_1, \ldots, d_k)$, where $d_1, \ldots, d_k$ are nonnegative integers denoting the degrees of differencing of the components. Suppose that $d_j > 0$ for at least one j. Let $d = \max(d_1, \ldots, d_k)$ and consider first the case $d = 1$. For such a series, in addition to the outliers introduced in Tsay et al. (2000) we also entertain the multivariate ramp shift (MRS) defined by

$$Y_t = X_t + wR_t^{(h)}$$

where $R_t^{(h)} = (I - B)^{-1} S_t^{(h)}$ with $S_t^{(h)}$ being a step-function at the time index h, i.e. $S_t^{(h)} = 1$ if $t \geq h$ and $=$ 0 otherwise. This outlier implies a slope change in the multivariate series and it may occur in an $I(1)$ series. It is not expected to happen in a stationary series because the series has no time slope. Consequently, for an MRS, we assume that it only applies to the components of Y_t with $d_j = 1$, that is, the size of the outlier satisfies $w_j = 0$ if $d_j = 0$.

The series X_t can be transformed into a stationary one by taking the first difference. This transformation affects the outlier model as follows. In the MIO case, $(I - B) Y_t = (I - B) X_t + \widetilde{\Psi}(B) w I_t^{(h)}$, where $\widetilde{\Psi}(B) = \nabla \Psi(B)$. Therefore, an MIO produces an MIO in the differenced series. In the MAO case, $(I - B) Y_t =$

$(I-B)\,X_t + w\left(I_t^{(h)} - I_{t-1}^{(h)}\right)$, producing two consecutive MAOs with the same size but opposite signs. In the MLS case, $(I-B)\,Y_t = (I-B)\,X_t + wI_t^{(h)}$, resulting in an MAO of the same size. In the MTC case, $(I-B)\,Y_t = (I-B)\,X_t + (I-B)\,(I-\delta B)^{-1}\,wI_t^{(h)} = (I-B)\,X_t + \zeta\,(B)\,wI_t^{(h)}$, where $\zeta\,(B) = 1 + \zeta_1 B + \zeta_2 B^2 + \ldots$ such that $\zeta_j = \delta^{j-1}\,(1-\delta)$. Thus, an MTC produces an MTC with decreasing coefficients ζ_j. In the MRS case, $(I-B)\,Y_t = (I-B)\,X_t + wS_t^{(h)}$, which produces an MLS of the same size.

Note that the results in section 4 can be easily extended to these outliers. For instance, it can be shown that the directions that maximize or minimize the kurtosis of the projected series under the presence of two consecutive MAOs with the same size but opposite signs are the direction of the outlier or the direction orthogonal to it. Therefore, in the $I(1)$ case, we propose a procedure similar to that of the stationary case for the first differenced series. The procedure consists of the following steps:

1. Take the first difference of Y_t. Check for MLS as in Section 5.2.1. All the level shifts detected in the differenced series are incorporated as ramp shifts in the original series and are estimated jointly with the model parameters. If any of the ramp shifts is not significant, it is removed from the model. We repeat this process until all the ramp shifts are significant. Finally, we obtain a series $Y_t^* = Y_t - \sum_{i=1}^{r_R} w_i R_t^{(h)}$ which is free of ramp shifts.

2. Take the first difference of Y_t^*. The series $(I-B)\,Y_t^*$ can be affected by the outlier as

$$(I-B)\,Y_t^* = (I-B)\,X_t + \eta\,(B)\,wI_t^{(h)}$$

where $\eta\,(B) = \widetilde{\Psi}\,(B)\,wI_t^{(h)}$ for an MIO, $\eta\,(B) = w\left(I_t^{(h)} - I_{t-1}^{(h)}\right)$ for an MAO, $\eta\,(B) = wI_t^{(h)}$ for an MLS and $\eta\,(B) = (I-B)\,(I-\delta B)^{-1}\,wI_t^{(h)}$ for an MTC. We then proceed as in section 5.2.2. All the outliers detected in the differenced series are incorporated by the corresponding effect in the original series and are estimated jointly with the model parameters. If any of the outliers is not significant, it is removed from the model. We repeat the process until all the outliers are significant.

Note that the prior procedure can be applied to cointegrated series. In this case ∇Y_t is overdifferenced, implying that its moving average component contains unit roots. Nevertheless, this is not a problem for the proposed procedure, because the directions of the outliers will be in general different from the directions of cointegration. In other words, if v is a vector obtained by maximizing or minimizing the kurtosis coefficient, then it is unlikely to be a cointegration vector, and $v'\nabla Y_t = \nabla\,(v'Y_t)$ is stationary and invertible because $v'Y_t$ is a nonstationary series. However, if the series are cointegrated, then the final estimation should be carried out using the error correction model of Engle and Granger (1987):

$$\nabla Y_t = C + D_1 \nabla Y_{t-1} + \cdots + D_{p-1} \nabla Y_{t-p+1} - \Pi Y_{t-1} + A_t.$$

Note that if v is the cointegration vector, then $v'Y_t$ is stationary and $\nabla v'Y_t$ is overdifferenced. Although no relationship is expected between the outlier directions and the cointegration vector, we have checked by Monte Carlo simulations that the probability of finding the cointegration relationship as a solution of the optimization algorithm is very low. Specifically, we have generated 10000 series from a vector AR(1) model with two components and a cointegration relationship and found the directions in (9). To compare the directions with the cointegration vector, we have calculated the absolute value of the cosine of the angle between these two directions. The average value of this cosine is 0.62 with variance 0.09. It is easy to show that if the angle has a uniform distribution in the interval $(0, \pi)$, the distribution of the cosine of the angle has expectation 0.63 and variance 0.09. Next, we repeated the same experiment with the same series but affected by outliers, level shifts or transitory changes and we obtained in every case that the mean of the angles between the direction found and the cointegrating direction is the one that exits between the direction of the outlier and the cointegration direction. Therefore, we conclude that there should be no confusion between the cointegration vectors and the directions that maximize or minimize the kurtosis coefficient of the projected series.

Consider next the case $d = 2$, i.e. the series are $I(2)$. Define a multivariate quadratic shift as follows:

$$Y_t = X_t + wQ_t^{(h)}$$

where $Q_t^{(h)} = (I - B)^{-1} R_t^{(h)}$. This outlier introduces a change in the quadratic trend of the multivariate series. The series X_t can be transformed into a stationary one by taking the second differences. Hence a multivariate quadratic shift is transformed into a multivariate level shift, a multivariate level shift is transformed into a multivariate additive outlier, and so on. A similar procedure as that proposed for the $I(1)$ case applies. In fact, the discussion can be generalized to handle outliers in a general I(d) series.

7 Simulations and Computational Results

In this section, we investigate the computational aspects of the proposed procedures. First, we obtain critical values for all the test statistics considered in the procedures. Second, we use various ways to compare the

Table 1: Models used in simulation study.

$k=2$			
Models	1	2	3
Φ	$\begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}$	—	$\begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}$
Θ	—	$\begin{pmatrix} -0.7 & 0 \\ -0.1 & -0.3 \end{pmatrix}$	$\begin{pmatrix} -0.7 & 0 \\ -0.1 & -0.3 \end{pmatrix}$
$k=3$			
Models	4	5	6
Φ	$\begin{pmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.4 & 0 \\ 0.6 & 0.2 & 0.5 \end{pmatrix}$	—	$\begin{pmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.4 & 0 \\ 0.6 & 0.2 & 0.5 \end{pmatrix}$
Θ	—	$\begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.3 & 0 \\ -0.7 & 0 & -0.5 \end{pmatrix}$	$\begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.3 & 0 \\ -0.7 & 0 & -0.5 \end{pmatrix}$

test statistics for level-shift detection. Finally, we conduct a simulation study to compare the power of the multivariate and projection test statistics. To save space, we only show the results for the stationary case.

7.1 Critical values

We consider six VARMA(p,q) models in the simulation. The number of components is either $k=2$ or $k=3$, and the parameter matrices used are given in Table 1. The constant term of the models is always the vector 1_k. The residual covariance matrix is the identity matrix.

The two autoregressive parameter matrices have eigenvalues 0.27 and 0.72, and 0.27, 0.5 and 0.72, respectively, while the moving-average parameter matrices have eigenvalues -0.3 and -0.7, and -0.3, -0.5 and -0.7, respectively. Using the six models, we generate critical values of the test statistics Λ_A, Λ_T and Λ_I of (13) and

$$D_L = \max_{1\le i\le 2k} \max_{1\le t\le n} \left|C_t^i\right|, \qquad \Lambda_L = \max_{1\le i\le 2k} \max_{1\le t\le n} \left|\lambda_{L,t}^i\right|. \tag{14}$$

The two statistics for detecting level shifts are included for comparison purpose.

The sample sizes used are $n = 100$, 200, and 500. For a given model and sample size, we generated 10,000 series and used the proposed procedures of the previous sections to compute the test statistics. If an autoregression is needed in the procedures, we use AIC to select the order. Table 2 summarizes the empirical critical values of the simulation. Based on the results, we recommend some critical values in Table 3 for practical use to detect multivariate outliers.

Table 2: Empirical critical values of the test statistics considered. These values are based on sample sizes n = 100, 200, and 500 and 10,000 realizations. M denotes the models in Table 1.

		95 %					99 %				
n	M	Λ_I	Λ_A	Λ_L	Λ_T	D_L	Λ_I	Λ_A	Λ_L	Λ_T	D_L
100	1	3.62	3.93	3.22	3.78	1.33	4.02	4.30	3.61	4.16	1.44
100	2	3.64	3.86	3.06	3.75	1.33	4.03	4.32	3.40	4.12	1.43
100	3	3.64	3.65	3.27	3.76	1.33	3.96	4.08	3.71	4.14	1.44
100	4	3.93	4.10	3.34	3.92	1.36	4.27	4.51	3.81	4.33	1.46
100	5	3.97	4.20	3.23	3.92	1.36	4.32	4.62	3.72	4.35	1.46
100	6	3.87	3.89	3.36	3.98	1.36	4.29	4.32	3.82	4.46	1.46
200	1	3.81	3.98	3.30	3.89	1.40	4.13	4.39	3.72	4.23	1.53
200	2	3.82	3.95	3.10	3.93	1.40	4.23	4.40	3.59	4.33	1.51
200	3	3.79	3.84	3.34	3.87	1.40	4.08	4.14	3.71	4.20	1.52
200	4	4.10	4.22	3.38	4.06	1.40	4.39	4.68	3.78	4.72	1.56
200	5	4.11	4.33	3.20	4.06	1.42	4.49	4.81	3.70	4.60	1.57
200	6	4.14	4.00	3.41	4.04	1.42	4.54	4.34	3.79	4.68	1.56
500	1	4.08	4.18	3.41	4.19	1.44	4.52	4.65	3.91	4.64	1.61
500	2	4.14	4.17	3.21	4.15	1.43	4.55	4.62	3.77	4.50	1.59
500	3	4.06	4.00	3.43	4.17	1.43	4.49	4.40	3.86	4.52	1.59
500	4	4.32	4.39	3.48	4.39	1.46	4.75	4.80	3.93	4.87	1.63
500	5	4.26	4.42	3.38	4.33	1.47	4.79	4.87	3.83	4.76	1.66
500	6	4.28	4.22	3.58	4.38	1.49	4.68	4.58	3.99	4.70	1.63

Table 3: Recommended critical values of the test statistics considered for sample size n = 100, 200 and 500.

		95 %			99 %		
n	k	Λ_I, Λ_A, Λ_T	Λ_L	D_L	Λ_I, Λ_A, Λ_T	Λ_L	D_L
100	2	3.7	3.2	1.3	4.1	3.6	1.4
	3	4.0	3.3	1.3	4.4	3.8	1.4
200	2	3.9	3.3	1.4	4.2	3.7	1.5
	3	4.1	3.4	1.4	4.6	3.8	1.5
500	2	4.1	3.4	1.4	4.5	3.8	1.6
	3	4.3	3.5	1.4	4.7	3.9	1.6

7.2 Comparison of various test statistics for detecting level shifts

Next, we compare the performance of multivariate, LR projection and cusum test statistics for level-shift detection. To this end, we use sample sizes $n = 100$ and 200, and three different outlier sizes, which are $w_L = 3 \times 1_k$, $4 \times 1_k$ and a random w_L. The direction of the random w_L was generated by (a) drawing a uniform [0,1] random variable u for each component and (b) defining $w_{L,i} = -1$, 0 or 1 if u is in the interval $(0, 1/3)$, $(1/3, 2/3)$ or $(2/3, 1)$. The resulting vector w_L was then normalized to have the same norm as $3 \times 1_k$.

For a given sample size and level shift, we generated 1000 series and computed the test statistic J_{max} in (5) for a level shift, the maximum projection statistic Λ_L in (14) and the maximum cusum statistic in (14) based on the proposed procedure. We compare the statistics with their respective critical values in Table 3 at the 5% significance level and tabulate the number of times a level shift is detected. The results are given in the first part of Table 4 (see columns $J_{\max}$, Λ_L and D_L). For all the models considered, the cusum test outperforms the other two, but all three tests seem to have decent power when the sample size is 200.

We also study the power of these three statistics in the presence of other outliers. Specifically, for each model, we generated 1000 series of size $n = 100$. Each series is contaminated by an innovative outlier at $h_I = 20$ with size $w_I = w \times 1_k$, an additive outlier at $h_A = 40$ with size $w_A = -w \times 1_k$, a transitory change at $h_T = 80$ with size $w_T = -w \times 1_k$, and a level shift at $h_L = 60$ with size $w_L = w \times 1_k$, where $w = 3$ or 4. A random vector w generated by the same method as before is also used as the size for all outliers. We compute and compare the three test statistics of level shift with their respective critical values in Table 3 at the 5% significance level. The power of these three statistics are given in the second part of Table 4 (see columns $J_{\max}$, Λ_L and D_L). All three tests are affected by the presence of other outliers, but similar to the case of a single level shift, the cusum test continues to outperform the other two test statistics. Furthermore, we measured the power loss of each test by

$$\text{loss}\,(i) = 1 - \frac{\text{power with outliers in model } i}{\text{power with no outliers in model } i},$$

and obtained the mean power loss of the three test statistics for the 6 models used with $w = 3$. The averaged loss for the multivariate statistic is 27.7%, that for the projection statistics is 17.1%, and that for the cusum test is 9.4%. Therefore, the multivariate and projection test statistics for level shift seem to be more susceptible to masking effects than the cusum test statistic.

Finally, it is important to study the type-I error of the three statistics in the presence of other outliers. We use a generating procedure similar to that of power study to conduct the simulation. However, for each

Table 4: Frequency of properly detecting a level shift for the multivariate, projection and cusum test statistics, where n is the sample size, M denotes the model in Table 1, h_i denotes time point at which a type i outlier occurs, and w is the outlier.

						$w = 3 \times 1_k$			$w = 4 \times 1_k$			$w = random$		
n	M	h_I	h_A	h_L	h_T	$J_{\max}$	Λ_L	D_L	$J_{\max}$	Λ_L	D_L	$J_{\max}$	Λ_L	D_L
100	1	—	—	50	—	70.0	83.0	100	96.6	98.2	100	68.9	83.9	100
100	2	—	—	50	—	58.3	82.6	100	89.2	96.9	100	67.6	94.0	100
100	3	—	—	50	—	46.5	73.3	87.6	91.5	93.2	100	66.9	81.6	95.3
100	4	—	—	50	—	93.6	92.7	99.6	100	98.8	100	83.6	87.6	99.7
100	5	—	—	50	—	68.1	98.8	100	94.7	99.7	100	76.9	91.3	100
100	6	—	—	50	—	86.3	86.6	88.7	98.9	99.0	99.2	78.2	77.2	96.7
200	1	—	—	100	—	80.7	95.4	100	98.1	99.6	100	90.3	91.0	100
200	2	—	—	100	—	92.2	95.6	100	97.0	99.3	100	88.3	98.0	100
200	3	—	—	100	—	78.1	90.6	99.5	97.1	98.4	100	80.9	84.6	100
200	4	—	—	100	—	98.7	98.9	100	100	100	100	90.6	94.0	100
200	5	—	—	100	—	85.8	99.8	100	99.0	100	100	89.3	98.7	100
200	6	—	—	100	—	97.5	97.8	100	99.6	99.7	100	90.3	83.6	100
100	1	20	40	60	80	45.6	76.0	92.6	73.0	91.8	99.3	55.2	70.9	95.7
100	2	20	40	60	80	52.0	87.6	100	74.3	96.7	100	58.9	77.6	99.7
100	3	20	40	60	80	17.0	53.6	66.6	40.3	84.3	86.6	44.1	49.2	80.9
100	4	20	40	60	80	73.6	74.0	90.3	92.0	85.0	100	67.9	73.9	94.0
100	5	20	40	60	80	61.3	81.6	100	71.3	92.0	100	65.9	82.9	99.7
100	6	20	40	60	80	63.6	55.3	74.3	87.0	80.0	95.0	58.2	49.5	85.3
100	1	25	50	—	75	4.0	14.0	3.3	2.6	26.3	3.3	3.0	21.0	5.0
100	2	25	50	—	75	0.3	9.0	3.0	1.3	11.0	2.0	1.3	17.4	5.3
100	3	25	50	—	75	4.0	21.3	4.3	2.3	31.3	4.6	9.0	41.4	6.3
100	4	25	50	—	75	6.3	15.6	3.0	7.0	28.0	6.3	10.4	24.4	6.0
100	5	25	50	—	75	3.3	8.0	4.0	2.3	12.6	4.3	4.3	15.4	5.0
100	6	25	50	—	75	16.3	22.3	4.3	16.6	27.3	4.6	15.4	37.8	4.7

Table 5: Empirical power of multivariate and projection test statistics for detecting an outlier in a vector time series, where n is the sample size, M denotes the model in Table 1, h is the time index of outlier and w is the size of the outlier.

				$w = 3 \times 1_k$		$w = 4 \times 1_k$		$w = random$	
	n	M	h	$J_{\max}$	Λ_I	$J_{\max}$	Λ_I	$J_{\max}$	Λ_I
MIO	100	1	50	59.9	77.8	95.0	98.0	63.2	80.6
	100	2	50	53.9	71.5	89.8	95.3	60.9	77.6
	100	3	50	51.1	68.6	88.0	95.2	47.5	76.6
	100	4	50	81.9	91.5	99.6	99.2	82.9	94.6
	100	5	50	62.0	76.1	95.1	97.4	73.6	86.6
	100	6	50	61.8	78.1	92.3	97.6	71.6	89.3
MIO	200	1	100	58.8	71.8	92.6	96.8	59.5	74.9
	200	2	100	58.5	68.4	92.5	95.5	56.9	72.9
	200	3	100	57.0	68.3	92.7	95.6	55.5	75.3
	200	4	100	81.6	87.5	100	100	80.6	91.3
	200	5	100	67.4	76.2	97.7	98.5	77.3	86.0
	200	6	100	67.1	75.8	98.2	98.8	78.6	86.0
	n	M	h	$J_{\max}$	Λ_A	$J_{\max}$	Λ_A	$J_{\max}$	Λ_A
MAO	100	1	50	86.6	93.6	99.3	99.3	80.6	84.6
	100	2	50	67.0	96.0	96.0	100	63.6	84.3
	100	3	50	91.0	99.0	99.6	100	87.3	95.7
	100	4	50	98.3	99.3	99.6	100	91.6	94.0
	100	5	50	78.6	95.6	98.6	100	78.6	92.3
	100	6	50	97.0	99.0	99.6	100	92.0	92.3
MAO	200	1	100	87.3	93.6	99.3	100	71.9	75.6
	200	2	100	67.0	91.0	95.6	99.3	64.6	83.3
	200	3	100	98.0	98.0	100	100	92.6	95.7
	200	4	100	99.3	99.6	100	100	97.3	93.6
	200	5	100	82.6	89.6	99.0	99.0	85.0	89.6
	200	6	100	98.8	97.8	100	100	97.0	95.6
	n	M	h	$J_{\max}$	Λ_T	$J_{\max}$	Λ_T	$J_{\max}$	Λ_T
MTC	100	1	50	61.3	88.6	93.3	98.6	60.9	97.7
	100	2	50	64.5	97.0	94.0	100	59.5	97.0
	100	3	50	71.3	93.6	92.6	99.3	57.2	95.7
	100	4	50	90.0	98.0	100	100	86.0	98.0
	100	5	50	71.0	97.6	95.3	99.6	79.9	99.3
	100	6	50	82.6	95.6	97.0	98.6	76.9	97.3
MTC	200	1	100	61.0	88.6	98.0	98.0	68.6	89.3
	200	2	100	66.0	92.0	94.6	99.3	68.9	92.0
	200	3	100	73.6	90.3	97.7	98.6	75.9	84.0
	200	4	100	92.0	92.8	99.5	99.5	92.6	94.7
	200	5	100	75.3	93.3	99.3	100	88.0	96.7
	200	6	100	95.9	95.7	100	99.2	85.0	92.6

generated series, the outliers consist of (a) an innovational outlier at $h_I = 25$ with size $w_I = w \times 1_k$, (b) an additive outlier at $h_A = 50$ with size $w_A = -w \times 1_k$, and (c) a transitory change at $h_T = 75$ with size $w_T = w \times 1_k$, where $w = 3$ or 4. Again, we also used a random vector w generated as before for the size of all outliers. The last six rows of Table 4 give the frequencies that the test statistic is greater than its empirical 95th percentile of Table 3. These frequencies denote chances of a false detection of a level shift by the three statistics. Once again, the cusum statistic outperforms the other two in maintaining the size of a test. The multivariate and projection statistics seem not robust to the presence of other outliers.

7.3 Power comparison between the multivariate and univariate statistics

In this subsection, we investigate the power of the test statistics for detecting other types of outlier. The outliers considered are multivariate additive and innovational outliers and transitory change. Again, we used the six models in Table 1 and sample sizes $n = 100$ and 200. The outlier occurs at $t = n/2$ and assumes three possible sizes as before. For each combination of model, sample size, and outlier, we generated 1000 series to compute the proposed test statistics. We then compared the statistics with their empirical 95th percentiles of Table 3 and tabulated the frequencies of detecting a significant outlier. Table 5 summarizes the power of various test statistics. From the table, it seems that projection test statistics outperform their corresponding multivariate counterparts. Overall, our limited simulation study supports the use of projections and cusum statistics in detecting outliers in a vector time series.

8 An Illustrative Example

We illustrate the performance of the proposed procedures by analyzing a real data example. The data are the logarithms of the annual gross national product (GNP) of Spain, Italy and France from 1947 to 2003. The series have 57 observations and are shown by solid lines in Figure 2.

As the series are clearly nonstationary we take the first difference of each GNP series. We then compute the projection directions using the proposed procedure in section 5 and apply the level shift detection algorithm to detect ramp shifts in the original series. The critical value is 1.3 as shown in Table 3. The algorithm detects a change point at time $h_1^L = 1975$. The value of the test statistic (11) for the time index is 1.39. To estimate the effect of the ramp shift, we first check if the series are cointegrated using Johansen's test (Johansen, 1991). A cointegration vector β is found and we use AIC to select the following error correction model with a cointegrating vector

$$\nabla Y_t = D_1 \nabla Y_{t-1} - \alpha\beta' Y_{t-1} + A_t, \tag{15}$$

where the estimated parameters are

$$\widehat{D}_1 = \begin{pmatrix} 0.299 & 0.095 & 0.510 \\ 0.069 & 0.344 & 0.524 \\ 0.100 & 0.221 & 0.728 \end{pmatrix}, \qquad \widehat{\alpha} = \begin{pmatrix} 0.007 \\ \text{-}0.001 \\ 0.003 \end{pmatrix}, \qquad \widehat{\beta} = \begin{pmatrix} 10.762 \\ \text{-}22.355 \\ 11.975 \end{pmatrix}.$$

Note that (15) is equivalent to the VAR(2) model $Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + A_t$ with $\widehat{\Pi}_1 = I + \widehat{D}_1 - \widehat{\alpha}\widehat{\beta}'$ and $\widehat{\Pi}_2 = -\widehat{D}_1$. Second, using this model we remove the effect of the ramp shift by estimating the following regression model:

$$A_t = \left(I - \widehat{\Pi}_1 B - \widehat{\Pi}_2 B^2\right) wR_t^{(1975)} + E_t,$$

and the series free from the effect of the ramp shift is obtain by

$$Y_t^* = Y_t - \widehat{w}R_t^{(1975)}.$$

Next, we look for other outliers using a critical value 4 taken from Table 3. Table 6 summarizes the results of the detection procedure. It identifies an MLS in 1966. The identified outlier is estimated and its effects on the series removed as in the case of the MRS. The procedure then detects an MAO in 1975, which is estimated and cleaned from the series. The procedure fails to detect any other outliers and is terminated. The outlier-adjusted series are shown by dashed lines in Figure 2.

After identifying the outliers for the series, we estimate jointly the outlier effects and the model parameters using a first-order vector error correction model with a cointegration relationship. The estimated effects of the three detected outliers are given in Table 7 along with the t-ratios of the estimates. The table shows some characteristics of the proposed procedure. The ramp shift detected by the algorithm in 1975 means a recession in all three countries and can be associated with the first oil crisis. This ramp shift can also be seen from the plot of the series. The algorithm also identifies a multivariate additive outlier in 1975 affecting especially the GNPs of Italy and France. Note that the procedure allows for multiple outlier detections at a

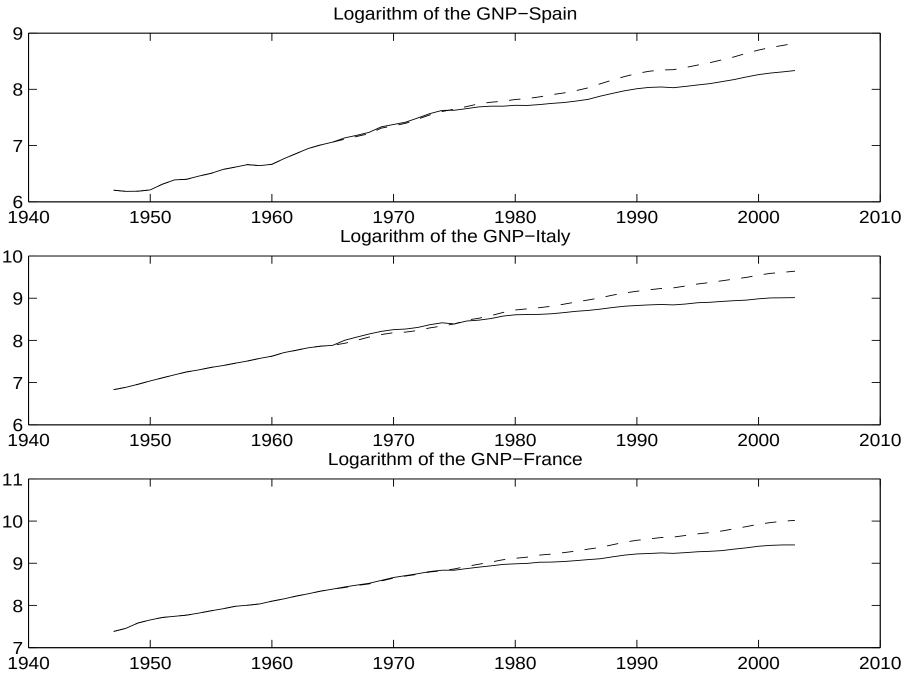

Figure 2: Original (solid lines) and Modified Logarithms of the GNP of Spain, Italy and France.

Table 6: Outliers found by the proposed algorithm.

					Outlier	
Iterations	(Λ_I, h_I)	(Λ_A, h_A)	(Λ_L, h_L)	(Λ_T, h_T)	Time	Type
1	(4.11,1966)	(4.05,1965)	(4.78,1966)	(4.22,1966)	1966	MLS
2	(3.37,1976)	(4.77,1975)	(4.15,1975)	(4.45,1975)	1975	MAO
3	(3.14,1960)	(3.74,1960)	(3.49,1960)	(3.68,1960)	—	—

time point. The final fitted vector error correction model is

$$\nabla Y_t = \begin{pmatrix} 0.2856 & 0.1839 & 0.3461 \\ 0.0341 & 0.6710 & 0.2721 \\ -0.0912 & 0.3765 & 0.5778 \end{pmatrix} \nabla Y_{t-1} - \begin{pmatrix} 0.007 \\ -0.000 \\ 0.002 \end{pmatrix} \begin{pmatrix} 14.412 & -21.651 & 8.132 \end{pmatrix} Y_{t-1} + A_t.$$

There are marked changes in the parameter estimates of the model with and without outlier detection. For instance, substantial changes in the diagonal elements of the D_1 matrix are observed before and after the outlier detection for the Italian and French GNP. The estimates of the cointegration vector also change. The estimated long-run equilibrium relationship between the variables before outlier detection was roughly

Table 7: Estimation of the sizes of the ouliers detected by the algorithm.

Time	Type	$\widehat{w}_1$ $(t\!-\!ratio)$	$\widehat{w}_2$ $(t\!-\!ratio)$	$\widehat{w}_3$ $(t\!-\!ratio)$
1966	MLS	0.0165 (1.7046)	0.0473 (7.0546)	0.0152 (2.0114)
1975	MRS	−0.0167 (−1.9723)	−0.0224 (−2.4668)	−0.0196 (−2.2817)
1975	MAO	−0.0434 (−1.8392)	−0.0672 (−4.1121)	−0.0312 (−1.6917)

Table 8: Summary of the procedure proposed in Tsay et al. (2000).

					Outlier	
Iterations	(J_I, h_I)	(J_A, h_A)	(J_L, h_L)	(J_T, h_T)	Time	Type
1	(15.08,1966)	(15.54,1965)	(11.39,1975)	(14.11,1966)	—	—
2	(3.78,1966)	(3.84,1965)	(3.16,1975)	(3.42,1966)	—	—

(.5S+.5F)-I, where S, F and I denote the log GNP of Spain, France and Italy, respectively. After outlier modeling, the cointegration vector roughly becomes (.64S+.36F)-I, which gives heavier weight to the Spanish GNP.

Finally, we compare the results with those obtained by applying the procedure of Tsay et al. (2000). The critical values for the multivariate statistics considered are 17.3 for MIO, MAO, and MTC, and 14.8 for MLS. The critical values for the component statistics are 3.9 for MIO, MAO and MTC and 3.6 for MLS. Table 8 summarizes the results using the same first-order vector error correction model. The procedure fails to detect any outliers.

9 Appendix: Proofs

Proof of Lemma 1. The kurtosis of y_t can be written as:

$$\gamma_y(v) = E\left[\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \frac{1}{n}\sum_{l=1}^{n} y_l\right)^4\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[\left(y_t - \frac{1}{n}\sum_{l=1}^{n} y_l\right)^4\right].$$

As $y_t = x_t + r_t^{(h,n)}$ and taking into account that x_t and $r_t^{(h,n)}$ are independent and $E[x_t] = E\left[x_t^3\right] = 0$, we get

$$E\left[\left(y_t - \frac{1}{n}\sum_{l=1}^{n} y_l\right)^4\right] = E\left[(x_t + r_t - \overline{r})^4\right] = E\left[x_t^4\right] + 6E\left[x_t^2\right](r_t - \overline{r})^2 + (r_t - \overline{r})^4,$$

and

$$\gamma_y(v) = \frac{1}{n}\sum_{t=1}^{n}\left(E\left[x_t^4\right] + 6E\left[x_t^2\right](r_t - \overline{r})^2 + (r_t - \overline{r})^4\right)$$
$$= \frac{1}{n}\sum_{t=1}^{n} E\left[x_t^4\right] + \frac{6}{n}\sum_{t=1}^{n} E\left[x_t^2\right](r_t - \overline{r})^2 + \frac{1}{n}\sum_{t=1}^{n}(r_t - \overline{r})^4.$$

Finally, as $E\left[x_t^2\right] = v'v$, $E\left[x_t^4\right] = 3E\left[x_t^2\right]^2 = 3(v'v)^2$, $\frac{1}{n}\sum_{t=1}^{n}(r_t - \overline{r})^2 = v'\Sigma_R v$ and $v'v = v'\Sigma_Y v - v'\Sigma_R v$, we obtain

$$\gamma_y(v) = 3(v'\Sigma_Y v)^2 - 3(v'\Sigma_R v)^2 + \omega_r(v).$$

■

Proof of Lemma 2. The Lagrangian for the extreme points of $\gamma_y(v)$ is

$$\pounds(v) = 3 - 3(v'\Sigma_R v)^2 + \omega_r(v) - \lambda(v'\Sigma_Y v - 1),$$

and its gradient is

$$\nabla\pounds(v) = -12(v'\Sigma_R v)\Sigma_R v + \left(\frac{4}{n}\sum_{t=1}^{n}(r_t - \overline{r})^2 B_t\right)v - 2\lambda\Sigma_Y v = 0.$$

Multiplying by v' in $\nabla\pounds(v)$ and taking into account the constraint $v'\Sigma_Y v = 1$, we have $\lambda = -6(v'\Sigma_R v)^2 + 2\omega_r(v)$. As $\Sigma_R = \frac{1}{n}\sum_{t=1}^{n} B_t$, then

$$-12(v'\Sigma_R v)\Sigma_R v + 4\left(\frac{1}{n}\sum_{t=1}^{n}(v'B_t v)B_t\right)v = \left(-12(v'\Sigma_R v)^2 + \frac{4}{n}\sum_{t=1}^{n}(v'B_t v)^2\right)(I + \Sigma_R)v.$$

Therefore,

$$-3\left(v'\Sigma_R v\right)\Sigma_R v+3\left(v'\Sigma_R v\right)^2\Sigma_R v+\left(\frac{1}{n}\sum_{t=1}^{n}\left(v'B_t v\right)B_t\right)v-\frac{1}{n}\sum_{t=1}^{n}\left(v'B_t v\right)^2\Sigma_R v$$
$$=-3\left(v'\Sigma_R v\right)^2 v+\frac{1}{n}\sum_{t=1}^{n}\left(v'B_t v\right)^2 v,$$

and, finally,

$$\sum_{t=1}^{n}\left[\left(v'B_t v\right)-3\left(v'\Sigma_R v\right)-\frac{\mu\left(v\right)}{n}\right]B_t v=n\left(v'\Sigma_R v\right)^2\left(\gamma_r\left(v\right)-3\right)v.$$

This implies that the extreme directions of $\mathcal{L}\left(v\right)$ under the constraint $v'\Sigma_Y v=1$ are the eigenvectors of the matrix

$$\left[\sum_{t=1}^{n}\beta_t\left(v\right)B_t\right]v=\mu\left(v\right)v,$$

where $\beta_t\left(v\right)=\left[\left(v'B_t v\right)-3\left(v'\Sigma_R v\right)-\frac{\mu(v)}{n}\right]$ and $\mu\left(v\right)=n\left(v'\Sigma_R v\right)^2\left(\gamma_r\left(v\right)-3\right)$. From (8), we get that:

$$\gamma_y\left(v\right)=3-\sigma_r^4\left(3-\gamma_r\left(v\right)\right)=3+\frac{\mu\left(v\right)}{n}.$$

Therefore, the maximum or the minimum of $\gamma_y\left(v\right)$ will be given when $\mu\left(v\right)$ is as large or as small as possible respectively, and the maximum and the minimum of the kurtosis will be given by the maximum and the minimum of the eigenvalues of the matrix $\sum_{t=1}^{n}\beta_t\left(v\right)B_t$. ■

Proof of Theorem 3. In the proof, we use the following equalities:

$$v'\Sigma_R v=\frac{1}{n}\sum_{t=1}^{n}\left(r_t-\overline{r}\right)^2,\quad v'B_t v=\left(r_t-\overline{r}\right)^2,\quad \left(v'\Sigma_R v\right)^2\gamma_r\left(v\right)=\frac{1}{n}\sum_{t=1}^{n}\left(r_t-\overline{r}\right)^4,$$

where $\overline{r}=\frac{1}{n}\sum_{t=1}^{n}r_t$.

1. In the MAO case, $r_h=v'w$, $r_t=0$, $\forall t\neq h$ and $\overline{r}=\frac{1}{n}r_h$. First, $n\left(v'\Sigma_R v\right)^2\gamma_r\left(v\right)=c_1 r_h^4$ and $v'\Sigma_R v=c_2 r_h^2$, where

$$c_1=\left(1-\frac{1}{n}\right)\left[\left(1-\frac{1}{n}\right)^3+\frac{1}{n^3}\right],\quad c_2=\frac{1}{n}\left(1-\frac{1}{n}\right),$$

and consequently, the eigenvalues are given by $\mu(v) = c_0 r_h^4$, where

$$c_0 = c_1 - 3nc_2^2 = \left(1 - \frac{1}{n}\right)\left[1 - \frac{6}{n}\left(1 - \frac{1}{n}\right)\right].$$

On the other hand, after some algebra it can be shown that

$$\left[\sum_{t=1}^{n} \beta_t(v) B_t\right] v = \left[m_1 r_h^3 + m_2 r_h^5\right] R_h,$$

where

$$m_1 = \left(1 - \frac{1}{n}\right)\left[\frac{1}{n^3} + \left(1 - \frac{1}{n}\right)^3 - 3c_2\right], \quad m_2 = -c_0 \frac{1}{n}\left(1 - \frac{1}{n}\right).$$

As $R_h = w$,

$$v = \frac{m_1 r_h^3 + m_2 r_h^5}{c_0 r_h^4} w,$$

and the other eigenvectors are orthogonal to w. Moreover, as the eigenvalues are given by $c_0 r_h^4$ and $c_0 > 0$ for $n > 5$, we get that the maximum of the kurtosis coefficient is given in the direction of w, while the minimum is attained in the orthogonal directions to w.

2. In the MTC case, $r_t = 0$ if $t < h$, $r_t = \delta^{t-h} r_h$ for $t \geq h$ and $\overline{r} = m r_h$, where $m = \left(1 - \delta^{n-h+1}\right) / (n(1-\delta))$. First, $n\left(v'\Sigma_R v\right)^2 \gamma_r(v) = c_1 r_h^4$ and $v'\Sigma_R v = c_2 r_h^2$, where

$$c_1 = (h-1) m^4 + \sum_{t=h}^{n} \left(\delta^{t-h} - m\right)^4, \quad c_2 = \frac{1}{n}\left[(h-1) m^2 + \sum_{t=h}^{n} \left(\delta^{t-h} - m\right)^2\right],$$

and consequently, the eigenvalues are given by $\mu(v) = c_0 r_h^4$, where $c_0 = c_1 - 3nc_2^2$. On the other hand, after some algebra, it can be shown that

$$\left[\sum_{t=1}^{n} \beta_t(v) B_t\right] v = \left[m_1 r_h^3 + m_2 r_h^5\right] R_h,$$

where

$$m_1 = (h-1)\left(m^4 - 3c_2 m^2\right) + \sum_{t=h}^{n} \left[\left(\delta^{t-h} - m\right)^4 - 3c_2\left(\delta^{t-h} - m\right)^2\right]$$
$$m_2 = -\frac{c_0}{n}\left[(h-1) m^2 + \sum_{t=h}^{n} \left(\delta^{t-h} - m\right)^2\right],$$

and, consequently, one eigenvector is proportional to w and the others are orthogonal to it. As the eigenvalues are given by $c_0 r_h^4$, the kurtosis coefficient of y_t is maximized or minimized when v is proportional to w depending on the sign of c_0, that in general depends on the values of n, h and δ.

3. In the MLS case, $r_t = 0$ if $t < h$, $r_t = r_h$ for $t \geq h$ and $\overline{r} = \frac{n-h+1}{n} r_h$. First, $n \left(v' \Sigma_R v\right)^2 \gamma_r(v) = c_1 r_h^4$ and $v' \Sigma_R v = c_2 r_h^2$, where

$$c_1 = \frac{(h-1)(n-h+1)}{n^4} \left[(n-h+1)^3 + (h-1)^3 \right], \quad c_2 = \frac{(h-1)(n-h+1)}{n^2},$$

and consequently, the eigenvalues are given by $\mu(v) = c_0 r_h^4$, where

$$c_0 = c_1 - 3nc_2^2 = \frac{(h-1)(n-h+1)}{n^3} \left[n^2 - 6n(h-1) + 6(h-1)^2 \right].$$

On the other hand, after some algebra, it can be shown that

$$\left[\sum_{t=1}^{n} \beta_t(v) B_t \right] v = \left[m_1 r_h^3 + m_2 r_h^5 \right] R_h,$$

where

$$m_1 = \frac{(h-1)(n-h+1)}{n^4} \left[(n-h+1)^3 + (h-1)^3 - 3c_2 \right] \qquad m_2 = -c_0 \frac{(h-1)(n-h+1)}{n^2}$$

showing that one eigenvector is proportional to w and the others are orthogonal to it. The eigenvalues are given by $c_0 r_h^4$ and it is not difficult to see that

$$\begin{array}{lcl} c_0 < 0 & \Longleftrightarrow & h \in \left(1 + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) n, 1 + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) n \right) \\ c_0 > 0 & \Longleftrightarrow & h \notin \left(1 + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) n, 1 + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) n \right). \end{array}$$

Therefore the maximum of the kurtosis coefficient is given in the direction of w if $c_0 > 0$ and the minimum of the kurtosis coefficient is given in the direction of w if $c_0 < 0$.

■

Proof of Corollary 4. The result follows immediately from Theorem 3, because the relation $A_t = E_t + w I_t^{(h)}$ coincides with the MAO case in a white noise series. ■

REFERENCES

Bai, J. (1994) "Least Squares Estimation of a Shift in linear processes", *Journal of Time Series Analysis,* 15, 453-472.

Balke, N. S. (1993) "Detecting level shifts in time series", *Journal of Business and Economic Statistics,* 11, 81-92.

Bianco, A. M., Garcia Ben, M., Martínez, E. J. and Yohai, V. J. (2001) "Outlier Detection in Regression Models with ARIMA Errors using Robust estimates", *Journal of Forecasting*, 20, 565-579.

Billingsley, P. (1968) *Convergence of Probability Measures*, John Wiley & Sons.

Carnero, M. A., Peña, D. and Ruiz, E. (2003) "Detecting level shifts in the presence of conditional heterokedasticity" Technical Report, Universidad Carlos III de Madrid.

Chang I. and Tiao G. C. (1983) "Estimation of Time Series Parameters in the Presence of Outliers" Technical Report 8, Statistics Research Center, University of Chicago.

Chang I., Tiao, G. C. and Chen, C. (1988) "Estimation of Time Series Parameters in the Presence of Outliers", *Technometrics,* 3, 193-204.

Chen, C. and Liu, L. (1993) "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association*, 88, 284-297.

Engle, R. and Granger, C. W. J. (1987) "Co-integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, 55, 251-276.

Fox, A. J. (1972) "Outliers in Time Series", *Journal of the Royal Statistical Society B,* 34, 350-363.

Huber, P. (1985) "Projection Pursuit (with discussion)", *The Annals of Statistics,* 13, 435-525.

Inclán, C. and Tiao, G. C. (1994) "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance", *Journal of the American Statistical Association,* 89, 913-923.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica,* 59, 1551-1580.

Jones, M. C. and Sibson, R. (1987) "What is Projection Pursuit (with discussion)?", *Journal of the Royal Statistical Society A,* 150, 1-36.

Justel, A., Peña, D. and Tsay, R. S. (2000) "Detection of Outlier Patches in Autoregressive Time Series", *Statistica Sinica,* 11, 651-673.

Le, N. D., Martin, R. D. and Raftery, A. E. (1996) "Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models", *Journal of the American Statistical Association,* 91, 1504-1515.

Luceño, A. (1998) "Detecting possibly non-consecutive outliers in industrial time series", *Journal of the Royal Statistical Society B,* 60, 295-310.

Lütkepohl, H. (1993) *Introduction to Multiple Time Series Analysis*, 2nd Ed., New York: Springer-Verlag.

McCulloch, R. E. and Tsay, R. S. (1993). "Bayesian inference and prediction for mean and variance shifts in autoregressive time series", *Journal of the American Statistical Association*, 88, 968–978.

McCulloch, R. E. and Tsay, R. S. (1994). "Bayesian analysis of autoregressive time series via the Gibbs sampler" *Journal of Time Series Analysis*, 15, 235–250.

Maravall, A. and Mathis, A. (1994) "Encompassing univariate models in multivariate time series", *Journal of Econometrics,* 61, 197-233.

Peña, D. and Prieto, F. J. (2001, *a*) "Multivariate Outlier Detection and Robust Covariance Matrix Estimation (with discussion)", *Technometrics,* 43, 286-310.

Peña, D. and Prieto, F. J. (2001, *b*) "Cluster Identification Using Projections", *Journal of the American Statistical Association,* 96, 1433-1445.

Posse, C. (1995) "Tools for two-dimensional exploratory projection pursuit", *Journal of Computational and Graphics Statistics,* 4, 83-100.

Priestley, M. B. (1981) *Spectral Analysis and Time Series*, London: Academic Press.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons.

Sánchez, M. J. and Peña, D. (2003), "The identification of Multiple Outliers in ARIMA models", *Communications in Statistics: Theory and Methods,* 32, 1265-1287.

Tsay, R. S. (1986) "Time Series Model Specification in the Presence of Outliers", *Journal of the American Statistical Association*, 81, 132-141.

Tsay, R. S. (1988) "Outliers, level shifts and variance changes in time series", *Journal of Forecasting*, 7, 1-20.

Tsay, R. S., Peña, D. and Pankratz, A.E. (2000) "Outliers in Multivariate Time Series", *Biometrika,* 87, 789-804.